# COSIGN: A Complete SDN Enabled All-Optical Architecture for Data Centre Virtualisation with Time and Space Multiplexing

C. Jackson [1] , K. Kondepu[1], Y. Ou[1], A. Beldachi [1], A. Pagès Cruz [2], F. Agraz [2], F. Moscatelli [3], W. Miao [4], V. Kamchevska [5], N. Calabretta [4], G. Landi [3], S. Spadaro[2] , R. Nejabati [1], D. Simeonidou [1]

[1] University of Bristol, United Kingdom, C.Jackson@Bristol.ac.uk
[2] Universitat Politècnica de Catalunya, Barcelona, Spain [3] Nextworks, Pisa, Italy [4] Technische Universiteit Eindhoven, Netherlands [5] Danmarks Tekniske Universiteit, Copenhagen, Denmark

**Abstract** *We present a Data Centre Virtualisation architecture with an SDN-controlled all-optical data plane combining OCS and TSON. Orchestration dynamically translates and provisions Virtual Data Centres requests onto the optical layer. We describe an implementation and characterisation of the data plane.*

## Introduction

Virtual Data Centres (VDC) provide highly configurable and scalable Infrastructure as a Service (IaaS) compute, network and storage resources. The underlying physical networks will require low latency, high bandwidth and high granularity and connectivity in order to efficiently serve the needs of many simultaneous tenants. These tenants have heterogeneous bandwidth requirements. Optical Time Division Multiplexing (TDM) or Time-Shared Optical Network (TSON) enables high granularity of bandwidth division. An additional requirement is a high degree of flexibility in order to support multi-tenancy. This is enabled using SDN control of all network elements. We propose and experimentally assess a novel 'complete' architecture. This paper gives the first detailed description of the implementation of a fully realised software stack ranging from the Orchestrator to the OpenFlow agents that enable the provisioning of VDC instances over the novel optical data layer containing TSON and optical circuit elements[1]. The intelligent orchestration layer dynamically provisions TDM slices or optical circuits to meet heterogeneous VDC request bandwidth requirements.

## VDC provisioning: Overall Architecture

The demonstration platform consists of compute server nodes and one control node. Each compute rack consists of three servers each with an OpenStack compute VM and is interfaced to both a TDM-based optical TOR device and a high-radix Polatis Large Port-Count Fibre Switch (LPFS) switch. We refer to these as the TSON and OCS networks respectively. The control node hosts the OpenStack control and SDN controller VMs. It is connected to the compute nodes via an Electronic Packet Switched (EPS) management network. The two optical data networks enable flexible optical network provisioning to meet a variety of requirements.

## Data Plane

The Data Plane consists of a LPFS providing Optical Circuit Switching (OCS), above which the beam-steering 2x2 4-core Multi-Core Fibre (MCF) switch[2] sits. This MCF device offers a

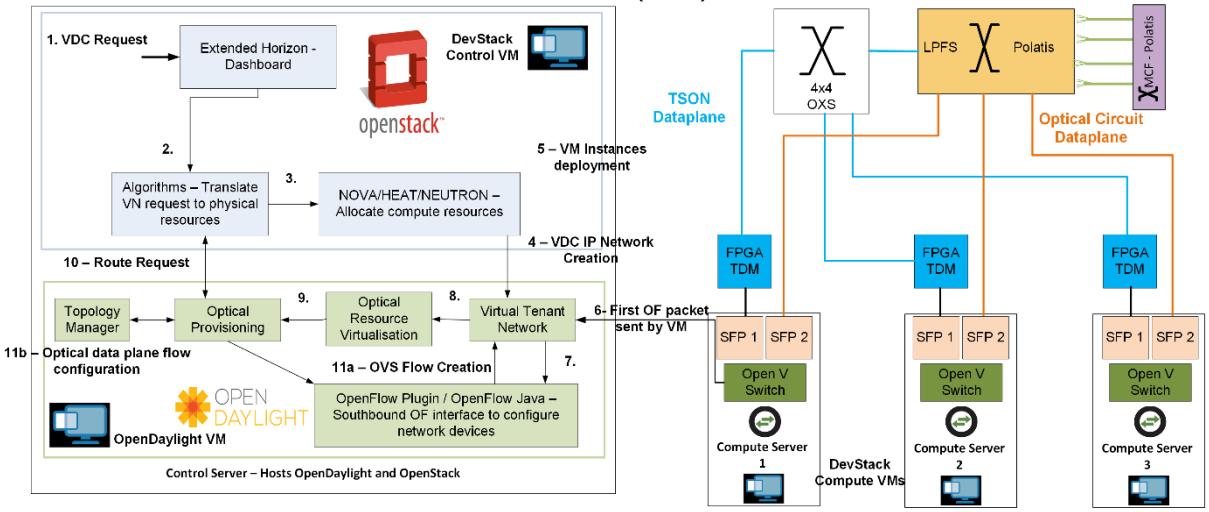


**Fig. 1**: Architecture and control flow for VDC provisioning

300% increase in fibre capacity over single-mode fibre (SMF) and we envision usage for inter-DC traffic. The two layers of optical circuit switching, combining SMF and MCF devices, each controlled via SDN enables a flat architecture and multi-dimensional optical switching. Additionally, there is a 4x4 optical fast switch (OXS) supporting broadcast with an FPGA controller that enables TDM switching and serves as a TOR. Compute and storage nodes are interfaced to optical TDM using FPGA TDM NIC cards[3]. We characterise the performance of the TDM network later. The novelty of the data plane lies in the combination of fully SDN-controlled and orchestrated TSON and OCS enabling granular bandwidth provisioning from the orchestration layer. Alongside MCF switching for inter-DC connectivity there is a range of data plane components to meet varied data path requirements.

**Compute Node Implementation**

Each compute or storage node in the architecture has a dedicated TDM connection to each node in the same rack. Inter-rack or high bandwidth (>8.6Gb/s) or low-latency communication is available through the LPFS backplane. Fig. 2(b) illustrates how the orchestration and control layer uses OVS to route flows to the provisioned interface. This arrangement also creates the possibility to apply dynamic circuit provisioning to adapt to bursts in traffic[4].

**SDN & Orchestration**

OpenDaylight Lithium (ODL) was enhanced with a number of novel extensions to support communication with the optical data plane elements. The OpenFlow protocol was extended to support the TSON and OCS devices. This included the definition and modelling of a TDM scheme, as well as the development of TSON OF agents for the OXS and FPGA NIC. The protocol extensions enabled provisioning of optical resources in combination with the orchestration layer. The orchestrator is an extended OpenStack (OSK) platform. For the optimized provisioning of optical resources, a novel algorithms module is used to translate tenants' bandwidth requirements into a request for TSON slots or an optical circuit, see Fig. 2(a). This is an extended version of a previously characterised data centre infrastructure[5]. The extended Northbound REST interface of ODL is used to interact with Optical Resource and Provisioning Modules to create the flow necessary to allocate the requested Virtual Network (VN). This is the first OpenStack (DevStack) implementation to dynamically provision TSON and OCS resources via an extended and optically-enabled SDN controller in this way. On each OSK compute node an OpenVswitch (OVS) is programmed to control flows between VM instances. Additionally, the new algorithms module developed specifically for this architecture determines the several logical instances (IP network, sub-network and ports) necessary to enable traffic exchange along the VDC instance. To map the VMs and create the logical resources, it interacts with the core orchestrator services via the OpenStack Heat service. As well as the physical route and the necessary timeslots, it also determines the particular VLAN to be employed when encapsulating the traffic of each virtual link, See Fig. 2(a) and 2(b).

**TDM Scheme**

The TDM scheme has configurable timeslot periods, 1-10µs, with the number of timeslots configurable from 4-96. We selected a 10µs frame with 96 slots. When a VDC is configured the bandwidth requirements between VMs are translated to physical requirements used to configure the OXS input to output mapping and the timeslots on which a TDM NIC may transmit. The compute OVS instances map VM source and destination MAC addresses to VLANs so that each NIC can divide the allocated timeslots between different VNCs, see Fig. 2(c). When data, uniquely identified by the Source-Destination MAC addresses enters the OVS at the source compute node, a VLAN ID is added to the packet header. When this data arrives at the destination compute node then the VLAN ID is

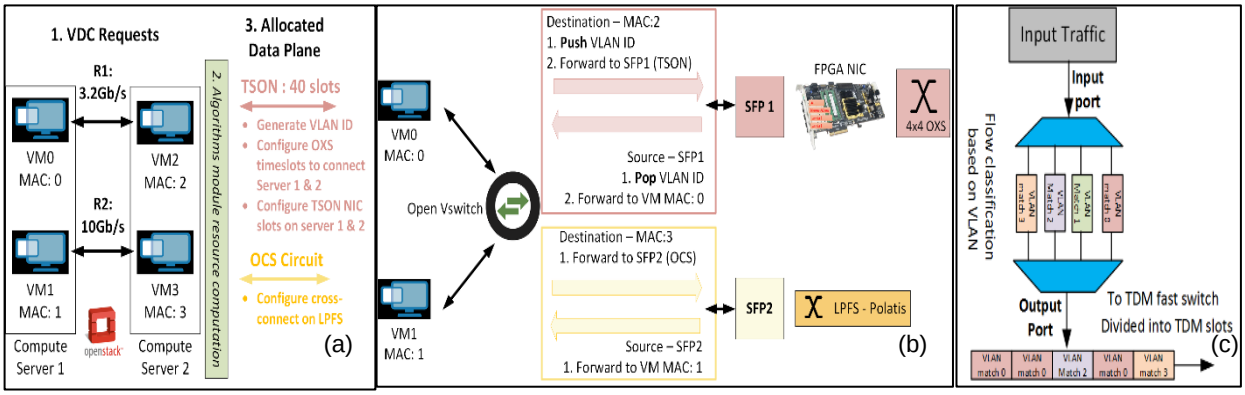


**Fig. 2**: (a) - Virtual network requests translated to TSON or OCS resources. (b) - OvS using VLAN push/pop actions to correctly route for the requests shown in (a). (c) - The TSON NIC using VLAN for flow control to allocate time slots

stripped, see Fig. 2. The flows to carry out these actions are calculated by the orchestrator and programmed via ODL. The 96 timeslots granularity in this configuration allows the provisioning of virtual connections from 84Mb/s to 8.6Gb/s (see Fig. 3).

**Experimental results**

We present a selection of results taken from the demonstration platform. The performance of the TDM network was measured in terms of throughput and latency against allocated timeslots. These results demonstrate a sustainable maximum data rate of up to 8.6Gbps, see Fig. 3 and Fig. 4. As can be observed better throughput and latency can be achieved with

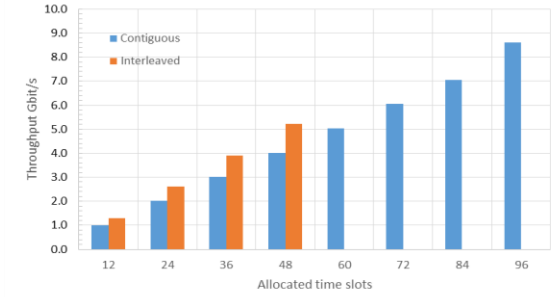


**Fig. 3**: Contiguous and interleaved allocated time slots vs. throughput for the TSON data plane

interleaved (or distributed) slots allocations. This is because interleaving reduces the maximum

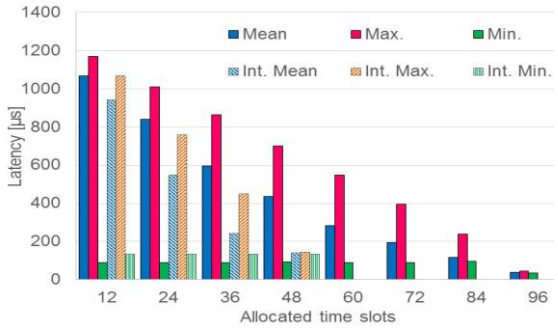


**Fig. 4**: TDM timeslot allocations against latency for contiguous and interleaved (Int.) slot allocation

delay between data transmissions. It is therefore recommended to avoid contiguous allocation for best performance. Similarly, in Fig. 4 the max. and mean latency measurements converge as timeslots increases because the largest gap between transmission slots reduces. The interleaved minimum is greater because unlike contiguous, there is always a no-transmit slot between transmissions.

The switching latency of the OXS was measured at both the circuit and application level. The time taken to switch an input from one output to another was measured electronically as 25*ns* at the circuit level. Using a ping-flood method we attempted to determine the effective reconfiguration time from an application perspective. The mean value was measured over five reconfigurations includes the end-to-end buffering and serialisation for the TSON scheme at 38.7µs. Compared to the same measurement with the NIC in Ethernet mode of 8.3µs, we identify an overhead of 30.1µs when using the extra buffering, logic and negotiation (key characters) involved in the TSON implementation. A similar experiment measured the mean reconfiguration time of the MCF switch over several iterations as 121µs. These results demonstrate that circuit-level measurements alone are insufficient to understand the performance impact of reconfiguration on the software network stack.

## Conclusion

A complete software stack from orchestrator-to-OpenFlow agents to enable VDCs on an all-optical data layer is successfully demonstrated. Circuit and TSON switching is combined to offer flexible and granular bandwidth provisioning. Our results allow us to make recommendations for provisioning TDM connections with different performance characteristics. Inter-DC traffic or very high bandwidth inter-cluster traffic is supported with a MCF switch connected to the high-radix LPFS backplane. We also compare circuit-level switch reconfiguration time with that observed from the application level and notice a significant increase. This demonstrates that signal-level measurements alone are insufficient for VDC architects and providers to fully understand the impact on application performance of optical switch reconfiguration times. The measured reconfiguration times at these different levels differ by orders of magnitude.

## Acknowledgements

This work is supported by ECFP7 grant no. 619572, the COSIGN project (www.fp7-cosign.eu).

## References

[1] C. R. Jackson et al., "Demonstration of the Benefits of SDN technology for All-Optical Data Centre Virtualisation," Proc. OFC (2017).

[2] H. C. H. Mulvad et al., "Beam-Steering All-Optical Switch for Multi-Core Fibers," Proc. OFC (2017).

[3] S. Yan *et al*., "Archon: A Function Programmable Optical Interconnect Architecture for Transparent Intra and Inter Data Center SDM/TDM/WDM Networking," in JLT (2015).

[4] Y. Ben-Itzhak et al., "C-Share: Optical Circuits Sharing for Software-Defined Data-Centers", Proc. Hotnets (2016).

[5] F. M. A. Pages et al., "Experimental Assessment of VDC Provisioning in SDN/OpenStack-based DC Infrastructures with Optical DCN," Proc. ECOC (2016).